UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/636,176	08/07/2003	Xuedong D. Huang	M61.12-0516	2089

27366 7590 08/24/2007
WESTMAN CHAMPLIN (MICROSOFT CORPORATION)
SUITE 1400
900 SECOND AVENUE SOUTH
MINNEAPOLIS, MN 55402-3319

EXAMINER	
LENNOX, NATALIE	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
08/24/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/636,176	HUANG ET AL.	
	Examiner	Art Unit	
	Natalie Lennox	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on 29 May 2007.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) 1-41 is/are pending in the application.

4a) Of the above claim(s) 2, 4-9, 31-33, and 41 is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) 1, 3, 10-30, and 34-40 is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date See Continuation Sheet.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

Continuation of Attachment(s) 3). Information Disclosure Statement(s) (PTO/SB/08), Paper No(s)/Mail Date :02/21/2007, 03/26/2007, 05/29/2007, 07/19/2007, 08/08/2007.

DETAILED ACTION

This Office Action has been issued in response to the amendments filed on May 29, 2007. Claims 1-41 are pending with claims 1, 3, 21, 29, 34, and 37 amended, and claims 2, 4-9, 31-33, and 41 cancelled.

Response to Arguments

1. Applicant's arguments with respect to claims 1, 12, 21, 29, 34, and 37 have been considered but are moot in view of the new ground(s) of rejection.

Claim Rejections - 35 USC § 103

2. The text of those sections of Title 35, U.S. Code not included in this action can be found in a prior Office action.

3. Claims 1, 3, 10, and 11 are rejected under 35 U.S.C. 103(a) as being unpatentable over Marshall (United States Statutory Invention Registration H1497) in view of Boesen (US Patent 6,094,492).

As per claim 1, Marshall teaches a headset comprising:

a head mount (headset 100 in Fig. 1 with head mount, also Fig. 3);

an audio microphone mechanically connected to the head mount (microphone 104 in Fig. 1, also present in Fig. 3); and

at least one earphone speaker mechanically connected to the head mount (Figs. 1 and 3 display two headsets each with two earphones), but Marshall does not specifically mention the headset comprising:

an in-ear transducer, configured to generate an electrical signal based on an input indicative of speech, and positioned to he located inside a user's ear and mechanically connected to the head mount.

However, Boesen teaches an in-ear transducer, configured to generate an electrical signal based on an input indicative of speech, and positioned to he located inside a user's ear and mechanically connected to the head mount (microphone 16 and ear-attachment portion 20 from Fig. 2, and also Col. 3, lines 58-64). It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of an in-ear transducer, configured to generate an electrical signal based on an input indicative of speech, and positioned to be located inside a user's ear and mechanically connected to the head mount as taught by Boesen for Marshall's headset because Boesen provides a voice sound transmitting unit using bone conduction and air conduction to obtain a pure voice sound signal for transmission minimizing interference from the surrounding sound environment (Col. 1, lines 7-12). It would have also been obvious to one of ordinary skill that the ear attachment portion could be attached to a headmount without affecting the proper position of the voice sound transmitting unit and providing the user with the option of having other devices attached to the headmount, such as another microphone, in order to obtain more information from the user and therefore make the system more reliable.

As per claim 3, Marshall, in view of Boesen, teach the headset according to claim 1, wherein the transducer comprises a microphone (Boesen's microphone 16 from Fig. 2, and Col. 3, lines 58-61).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of the transducer comprising a microphone as taught by Boesen for Marshall's headset because Boesen provides a voice sound transmitting unit using bone conduction and air conduction to obtain a pure voice sound signal for transmission minimizing interference from the surrounding sound environment (Col. 1, lines 7-12).

As per claim 10, Marshall, in view of Boesen, teach the headset of claim 1, wherein the transducer is rigidly connected to the head mount (Boesen's ear attachment portion 20 from Fig. 1, and also Col. 3 line 64 to Col. 4 line 3).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of the transducer rigidly connected to the head mount as taught by Boesen for Marshall's headset because provides a voice sound transmitting unit using bone conduction and air conduction to obtain a pure voice sound signal for transmission minimizing interference from the surrounding sound environment (Col. 1, lines 7-12). It would have also been obvious to one of ordinary skill that the ear attachment portion could be attached to a headmount without affecting the proper position of the voice sound transmitting unit and providing the user with the option of having other devices attached to the headmount, such as another microphone,

in order to obtain more information from the user and therefore make the system more reliable.

As per claim 11, Marshall, in view of Boesen, teach the headset of claim 10, wherein the audio microphone is rigidly connected to the head mount (Marshall's microphone 104 in Figs. 1 and 3, as shown connected to head mount).

As per claim 34, Marshall teaches an audio input system, comprising:

a headset including an audio microphone (headset 100 and microphone 104 from Fig.), but Marshall does not specifically mention the system comprising;

a speaker and an in-ear sensor configured to sense vibration in a user's ear and output a sensor signal indicative of the vibration.

However, Boesen teaches a speaker and an in-ear sensor configured to sense vibration in a user's ear and output a sensor signal indicative of the vibration (Col. 3, lines 5-7, also bone conduction sensor 14 from Fig. 2 with Col. 4, lines 26-35).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of an in-ear speech sensor configured to sense vibration within a user's ear and output a sensor signal indicative of the vibration as taught by Boesen for Marshall's system because Boesen provides a bone conduction voice transmission apparatus and system that transmits voice sound using bone conduction and air conduction to obtain a pure voice sound signal for transmission minimizing interference from the surrounding sound environment (Col. 1, lines 6-12), and also provides a speaker and receiver in the voice sound transmitting unit to enable a two-way communication (Col. 3, lines 5-7).

As per claim 35, Marshall, as modified by Boesen, teaches the audio input system of claim 34, wherein the audio microphone is configured to output a microphone signal based on a received audio input (Col. 2, lines 23-26, *information obtained from the microphone is used primarily to display "acoustical" data about the subject,* also Col. 3, lines 52-54, *analog signal apparatus 106 prepares the photo, infrared, and audio signals to be digitized by the analog-to-digital converter board*, signal apparatus 106 appears in both Fig. 1 and Fig. 2, where the audio signal outputted from microphone 104 enters component 106).

4. Claims 12-15, and 21-23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Holzrichter (US Patent 6,006,175) in view of Miyazawa et al. (US Patent 5,983,186).

As per claim 12, Marshall teaches a speech detection system comprising:

an audio microphone outputting a microphone signal based on an audio input (microphone 70 in Fig. 19, also Col. 31, lines 15-17);

a speech sensor configured to sense movement of a user's face and output a sensor signal indicative of the movement (tongue 21 and jaw 22 motion EM sensors in Fig. 4, also Col. 32, lines 31-35); and

a speech detector component configured to receive the sensor signal and output a speech detection signal indicative of whether the user is speaking based on the sensor signal (Col. 26, lines 34-38, lines 34-38, where combiner 67 in Fig. 12 contains *algorithmic decision tree joining one nonacoustic speech recognition (NASR) and one*

conventional acoustic speech recognition (CASR) algorithm; also Col. 25 line 63 to Col. 26 line 16 describes the process signals undergo until formed into feature vectors 61 from NASR and 65 from CASR (in Fig. 12), and further processed and combined to determine if speech is present in order to apply a speech recognition algorithm); but Holzrichter does not specifically mention the system comprising:

to control power to a speech recognizer based on the speech detection signal. However, Miyazawa et al. teach controlling power to a speech recognizer based on the speech detection signal (Col. 2, lines 46-53).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of controlling power to a speech recognizer based on the speech detection signal as taught by Miyazawa et al. for Holzrichter's system because Miyazawa provides a voice-activated interactive speech recognition device and method that performs recognition operations only when a recognizable speech input is detected to minimize power consumption (Col. 2, lines 46-49).

As per claim 13, Holzrichter, in view of Miyazawa et al., teach the speech detection system of claim 12, wherein the speech detector component is configured to receive the microphone signal and provide the speech detection signal based on the sensor signal and the microphone signal (Holzrichter's Col. 26, lines 34- 38, where combiner 67 in Fig. 12 contains *algorithmic decision tree joining one nonacoustic speech recognition OVASR) and one conventional acoustic speech recognition (CASR) algorithm*; also Col. 25 line 63 to Col. 26 line 16 describes the process signals undergo

until formed into feature vectors 61 from NASR and 65 from CASR (in Fig. 12), and further processed and combined to determine speech recognition).

As per claim 14, Holzrichter, in view of Miyazawa et al., teach the speech detection system of claim 12, wherein the speech sensor comprises a radiation sensor configured to sense radiation reflected from the user's face (Holzrichter's Col. 49, lines 6-8, the *EM wave acoustic microphones detect acoustic vibrations of human tissue, using EM wave sensors*; also Col. 49, lines 15-22, *EM wave generating, transmitting and detecting system, including infrared or visible wave radar that can penetrate the first surface of the skin, as well as reflect from the first skin-air surface [...] this includes their use in radiating modes*).

As per claim 15, Holzrichter, in view of Miyazawa et al., teach the speech detection system of claim 14, wherein the radiation sensor comprises an infrared sensor (Holzrichter's Col. 49, lines 59-65, use *EM radiation, including visible and IR* (infrared) *spectral information* 15-22, *EM wave system including [...] infrared or visible wave radar that can penetrate the first surface of the skin, as well as reflect from the first skin-air surface [...] this includes their use in radiating modes.*).

As per claim 21, Holzrichter teaches a method of detecting whether a user is speaking, comprising:

providing a sensor signal indicative of sensed radiation reflected from the user's face (Holzrichter's Col. 48, lines 44-48, method of speech characterization that uses electromagnetic (EM) radiation scattered (i.e., reflected and/or attenuated) from human

speech organs in concert with acoustic speech output for the purpose of speech recognition, also antenna 53 in Fig. 12 receives the sensor signal);

detecting whether the user is speaking based on the sensor signal (Col. 49, lines 30-34, where information on the positions and presence or absence of speech organ interfaces is provided by measuring the time between transmitted and received EM signals); but Holzrichter does not specifically mention the method comprising:

controlling power to a speech recognizer based on whether the user is speaking. However, Miyazawa et al. teach controlling power to a speech recognizer based on whether the user is speaking (Col. 2, lines 46-53).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of controlling power to a speech recognizer based on the speech detection signal as taught by Miyazawa et al. for Holzrichter's method because Miyazawa provides a voice-activated interactive speech recognition device and method that performs recognition operations only when a recognizable speech input is detected (Col. 2, lines 46-49).

As per claim 22, Holzrichter, in view of Miyazawa et al., teach the method of claim 21, wherein providing a sensor signal comprises:

directing infrared radiation on the user's face (Holzrichter's Col. 49, lines 15-19, *the use of EM wave generating, transmitting and detecting system, including [...] infrared or visible wave radar that can [...] reflect from the first skin-air surface*); and

detecting infrared radiation reflecting from the user's face (Holzrichter's Col. 49, lines 15-19, *the use of EM wave generating, transmitting and detecting system,*

including [...] infrared or visible wave radar that can [...] reflect from the first skin-air surface).

As per claim 23, Holzrichter, in view of Miyazawa et al., teach the method of claim 22, wherein providing a sensor signal comprises:

generating the sensor signal as a radiation detection signal indicative of a measure of the detected infrared radiation (Holzrichter's Col. 49, lines 15-19, *the use of EM wave generating, transmitting and detecting system, including [...] infrared or visible wave radar that can [...] reflect from the first skin-air surface*, also Col. 49, lines 30-34, where *information on the positions and presence or absence of speech organ interfaces is provided by measuring the time between transmitted and received EM signals*).

5. Claims 29 and 30 are rejected under 35 U.S.C. 103(a) as being unpatentable over Holzrichter (US Patent 6,006,175) in view of Boesen (US Patent 6,094,492) and Miyazawa et al. (US Patent 5,983,186).

As per claim 29, Holzrichter teaches a speech recognition system, comprising:

a speech detector system comprising:

an audio microphone outputting a microphone signal based on an audio input (microphone 70 in Fig. 19, also Col. 31, lines 15-17, acoustic in formation from microphone is inputted into an acoustic speech sensor);

a speech detector component configured to receive the sensor signal and output a speech detection signal indicative of whether the user is speaking based on the sensor signal (Col. 26, lines 34-38, where combiner 67 in Fig. 12 contains algorithmic

decision tree joining one nonacoustic speech recognition (NASR) and one conventional acoustic speech recognition (CASR) algorithm; also Col. 25 line 63 to Col. 26 line 16 describes the process signals undergo until formed into feature vectors 61 from NASR and 65 from CASR (in Fig. 12), and further processed and combined to determine if speech is present in order to apply a speech recognition algorithm);

a background speech removal component providing a modified speech signal based on the speech detection signal and the microphone signal (processor 66 in Fig. 12, Col. 26, lines 4-7, processor can include gain setting, speaker normalization, time adjustment, background removal, comparison to data from previous frames, and other well known procedures); and

a speech recognition engine receiving the modified speech signal and recognizing speech represented by the modified speech signal (speech recognition algorithm 68 in Fig. 12, Col. 26 lines 13-16, the two feature vectors are further processed and combined and if the result is speech recognition, a speech recognition algorithm 68 is applied); but Holzrichter does not specifically mention the system comprising:

an in-ear speech sensor configured to sense vibration within a user's ear and output a sensor signal indicative of the vibration; and

the speech recognition engine being powered based on the speech detection signal.

However, Boesen teaches an in-ear speech sensor configured to sense vibration within a user's ear and output a sensor signal indicative of the vibration (bone conduction sensor 14 from Fig. 2, and also Col. 4, lines 26-35).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of an in-ear speech sensor configured to sense vibration within a user's ear and output a sensor signal indicative of the vibration as taught by Boesen for Holzrichter's system because Boesen provides a bone conduction voice transmission apparatus and system that transmits voice sound using bone conduction and air conduction to obtain a pure voice sound signal for transmission minimizing interference from the surrounding sound environment (Col. 1, lines 6-12). Further, Holzrichter in view of Boesen, do not specifically mention the system comprising:

the speech recognition engine being powered based on the speech detection signal. However, Miyazawa et al. teach the speech recognition engine being powered based on the speech detection signal (Col. 2, lines 46-53).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of controlling power to a speech recognizer based on the speech detection signal as taught by Miyazawa et al. for Holzrichter's system, in view of Boesen, because Miyazawa provides a voice-activated interactive speech recognition device and method that performs recognition operations only when a recognizable speech input is detected to minimize power consumption (Col. 2, lines 46-49).

As per claim 30, Holzrichter, in view of Boesen and Miyazawa et al., teach the speech recognition system of claim 29, wherein the speech detector component is configured to receive the microphone signal and provide the speech detection signal based on the sensor signal and the microphone signal (Holzrichter's Col. 26, lines 34-38, where combiner 67 in Fig. 12 contains *algorithmic decision tree joining one nonacoustic speech recognition (NASR) and one conventional acoustic speech recognition (CASR) algorithm*; also Col. 25 line 63 to Col. 26 line 16 describes the process signals undergo until formed into feature vectors 61 from NASR and 65 from CASR (in Fig. 12), and further processed and combined to determine speech recognition).

6. Claim 16 is rejected under 35 U.S.C. 103(a) as being unpatentable over Holzrichter (US Patent 6,006,175) in view of Miyazawa et al. (US Patent 5,983,186) as applied to claim 14 above, and further in view of Bambot et al. (US Patent 6,590,651).

As per claim 16, Holzrichter, in view of Miyazawa et al., teach the speech detection system of claim 14, but doesn't specifically mention the radiation sensor comprising a charge coupled device. However, Bambot et al. teach an electromagnetic radiation sensor that may comprise a charge coupled device. (Col. 6, lines 44-47).

It would have been obvious to one of ordinary skill to use the feature of a charge coupled device as an electromagnetic radiation sensor as taught by Bambot et al. for Holzrichter's radiation sensor, in view of Miyazawa et al., because Bambot et al. provide a method and apparatus to irradiate a target tissue with electromagnetic radiation and to

detect returned electromagnetic radiation to determine a property, condition, or characteristics of the target tissue.

7. Claims 17 and 18 are rejected under 35 U.S.C. 103(a) as being unpatentable over Holzrichter (US Patent 6,006,175) in view of Miyazawa et al. (US Patent 5,983,186) as applied to claim 14 above, and further in view of Holzrichter et al. (US 2003/0097254).

As per claim 17, Holzrichter, in view of Miyazawa et al., teaches the speech detection system of claim 14, but doesn't specifically mention the speech detector component configured to detect a baseline value of a signal characteristic of the sensor signal.

However, Holzrichter et al. teach a speech segmentation procedure using threshold detection of an electromagnetic (EM) sensor signal to define onset and end of voiced segment timing (Paragraph [0031]).

It would have been obvious to one of ordinary skill in the art to have use the feature of threshold detection of an EM sensor signal as taught by Holzrichter et al. for Holzrichter's speech detector component, as modified by Miyazawa et al., because Holzrichter et al. provides a system for removing "excess" information from a human speech signal by using an EM sensor, a microphone, and their algorithms.

As per claim 18, Holzrichter, in view of Miyazawa et al., and further in view of Holzrichter et al., teach the speech detection system of claim 17, wherein the speech detection component configured to output the speech detection signal based on a value

of the signal characteristic during an observation time period relative to the baseline value (Holzrichter et al. in paragraph [0062], *the onset of speech event can be automatically determined by measuring a signal from the EM sensor that senses movement of a speech organ that reliably signals speech onset. [U]sing the EM sensor signal to measure the beginning of vocal fold movement and sending its signal to a processor* [that] *compares the measured glottal signal to a predetermined threshold level, which if it exceeds a predetermined threshold, defines a voiced speech onset time*; and also in paragraph [0063], *[I]f the EM sensor signal drops below a predetermined threshold signal, averaged over a predetermined time interval, the processor will note this time as an end of voiced speech segment time*).

It would have been obvious to one of ordinary skill in the art to have use the feature of the speech detection component configured to output the speech detection signal based on a value of the signal characteristic during an observation time period relative to the baseline value as taught by Holzrichter et al. for Holzrichter's speech detector component, as modified by Miyazawa et al., because Holzrichter et al. provides a system for removing "excess" information from a human speech signal by using an EM sensor, a microphone, and their algorithms.

1. Claim 19 is rejected under 35 U.S.C. 103(a) as being unpatentable over Holzrichter (US Patent 6,006,175) in view of Miyazawa et al. (US Patent 5,983,186) and Holzrichter et al. (US 2003/0097254) as applied to claim 18 above, and further in view of May, Jr. (US Patent 4,382,164).

As per claim 19, Holzrichter, as modified by Miyazawa et al. and Holzrichter et al., teach the speech detection system of claim 18, but they don't specifically mention the speech detector component to be configured to intermittently re-estimate the baseline value of the signal characteristic. May, Jr. teaches a threshold generator which develops dynamically adjustable decision levels for a speech definer in a speech detector. (Col. 4, lines 21-24, also threshold generator 14 in Fig. 1, which is a diagram of a basic speech detector).

It would have been obvious to one having ordinary skill in the art to have used the feature of an adjustable threshold as taught by May, Jr. for the speech detection system as taught by Holzrichter, as modified by Miyazawa et al. and Holzrichter et al., because May, Jr.'s invention relates to signal detecting arrangements for detecting speech activity in the presence of noise.

2. Claim 20 is rejected under 35 U.S.C. 103(a) as being unpatentable over Holzrichter (US Patent 6,006,175), in view of Miyazawa et al. (US Patent 5,983,186), as applied to claim 12 above, and further in view of Marshall (United States Statutory Invention Registration H1497).

As per claim 20, Holzrichter, as modified by Miyazawa et al., teaches the speech detection system of claim 12, but doesn't specifically mention the audio microphone and the speech sensor to be mounted to a headset. Marshall teaches a headset with an audio microphone and photo/thermal detectors (Fig. 1, headset 100, photodetectors and/or thermal detectors 102).

It would have been obvious to one having ordinary skill in the art to use the feature of mounting the audio microphone and speech detectors into a headset as taught by Marshall for Holzrichter's speech detection system, as modified by Miyazawa et al., because Marshall provides a combination of a photo/thermal detector and microphone for the purpose of conducting speech pronunciation measurements for detecting problems and help identify solutions relating to speech production for verbally challenged individuals in either the speech pathology, speech therapy, language learning, or basic education fields.

3. Claim 24 is rejected under 35 U.S.C. 103(a) as being unpatentable over Holzrichter (US Patent 6,006,175), in view of Miyazawa et al. (US Patent 5,983,186), as applied to claim 23 above, and further in view of May, Jr. (US Patent 4,382,164).

As per claim 24, Holzrichter, as modified by Miyazawa et al., teaches the method of claim 23, but doesn't disclose detecting whether the user is speaking comprising of intermittently calculating a baseline sensor signal value. May, Jr. teaches a threshold generator which develops dynamically adjustable decision levels for a speech definer in a speech detector. (Col. 4, lines 21-24, also threshold generator 14 in Fig. 1, which is a diagram of a basic speech detector).

It would have been obvious to one having ordinary skill in the art to have used the feature of an adjustable threshold as taught by May, Jr. for the speech detection system as taught by Holzrichter, as modified by Miyazawa et al. because May, Jr.'s

invention relates to signal detecting arrangements for detecting speech activity in the presence of noise.

8. Claims 25, 26, and 27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Holzrichter (US Patent 6,006,175), in view of Miyazawa et al. (US Patent 5,983,186) and May, Jr. (US Patent 4,382,164), as applied to claim 24 above, and further in view of Holzrichter et al. (US 2003/0097254).

As per claim 25, Holzrichter, as modified by Miyazawa et al. and May, Jr. teach the method of claim 24, but don't specifically mention detecting whether the user is speaking comprising comparing the sensor signal to the baseline sensor signal value. However, Holzrichter et al. teaches the onset of speech event can be automatically determined by measuring a signal from the EM sensor that senses movement of a speech organ that reliably signals speech onset. Using the EM sensor signal to measure the beginning of vocal fold movement and sending its signal to a processor [that] compares the measured glottal signal to a predetermined threshold level, which if it exceeds a predetermined threshold, defines a voiced speech onset time. If the EM sensor signal drops below a predetermined threshold signal, averaged over a predetermined time interval, the processor will note this time as an end of voiced speech segment time (paragraphs [0062] and [0063]).

It would have been obvious to one of ordinary skill in the art to have use the feature of threshold detection of an EM sensor signal as taught by Holzrichter et al. for Holzrichter's speech detector component, as modified above, because Holzrichter et al.

provides a system for removing "excess" information from a human speech signal by using an EM sensor, a microphone, and their algorithms.

As per claim 26, Holzrichter, as modified by Miyazawa et al., May, Jr., and Holzrichter et al., teach the method of claim 25, further comprising:

providing a microphone signal indicative of a sensed audio speech signal (Holzrichter's Col. 17, lines 24-27, *in Fig. 3, signals from acoustic microphone 1, and three EM sensors for vocal folds are combined using vocal tract model 5 to form a vocal tract feature vector 6*).

As per claim 27, Holzrichter, as modified by Miyazawa et al., May, Jr., and Holzrichter et al., teach the method of claim 26, wherein detecting whether the user is speaking comprises:

detecting whether the user is speaking based on the sensor signal and the microphone signal (Holzrichter's Col. 26, lines 34-38, where combiner 67 in Fig. 12 contains *algorithmic decision tree joining one nonacoustic speech recognition (NASR) and one conventional acoustic speech recognition (CASR) algorithm*; also Col. 25 line 63 to Col. 26 line 16 describes the process signals undergo until formed into feature vectors 61 from NASR and 65 from CASR (in Fig. 12), and further processed and combined to determine speech recognition).

9. Claim 28 is rejected under 35 U.S.C. 103(a) as being unpatentable over Holzrichter (US Patent 6,006,175) in view of Miyazawa et al. (US Patent 5,983,186) as applied to claim 21 above, and further in view of Nakamura (US Patent 4,769,845).

As per claim 28, Holzrichter, in view of Miyazawa et al., teaches the method of claim 21, but doesn't disclose providing a sensor signal comprising of sensing an image of the user's face and providing the sensor signal as an image signal indicative of the sensed image. Nakamura teaches an image pickup apparatus that picks up the image of the shape of a persons lip during speech as an optical image, and converts this optical image in a conventional manner into an electrical television image signal (Col. 2, lines 55-58).

It would have been obvious to one having ordinary skill in the art to use the feature of an image pickup apparatus as taught by Nakamura for the method of detecting whether a user is speaking as taught by Holzrichter and modified by Miyazawa et al., because Nakamura provides a method of recognizing speech from a lip image.

10. Claims 36-40 are rejected under 35 U.S.C. 103(a) as being unpatentable over Marshall (United States Statutory Invention Registration H1497), in view of Boesen (US Patent 6,094,492) as applied to claim 34 above, and further in view of Holzrichter (US Patent 6,006,175).

As per claim 36, Marshall, in view of Boesen, teaches the audio input system of claim 34, but doesn't disclose the system further comprising a speech detector component configured to receive the sensor signal and output a speech detection signal indicative of whether the user is speaking or is about to speak, based on the sensor signal. Holzrichter teaches a combiner that contains an algorithmic decision tree joining

one non-acoustic speech recognition (NASR) and one conventional acoustic speech recognition (CASR) algorithm, that after receiving the sensor and audio signals from the processor, combines them to determine if speech is present (combiner 67 in Fig. 12, also Col. 26, lines 34-38 and Col. 25 line 63 to Col. 26 line 16).

It would have been obvious to one having ordinary skill in the art to use the feature of a combiner as taught by Holzrichter for Marshall's audio input system, as modified by Boesen, because Holzrichter provides the use of non-acoustic information in combination with acoustic information for speech recognition and related speech technologies.

As per claim 37, Marshall, in view of Boesen, teaches a speech recognition system that comprises a headset including an audio microphone outputting a microphone signal based on an audio input (Marshall's headset 100, microphone 104, and signal apparatus 106 in Fig. 1, where the audio signal outputted from microphone 104 enters component 106, also Col. 2, lines 26-28, *the photo and infrared information is used to display "mouth/lips/tongue/motor" information about the subject*);

and an in-ear speech sensor configured to sense a physical characteristic indicative of speech and output a sensor signal indicative of the sensed physical characteristic (Boesen's bone conduction sensor 14 and air conduction sensor or microphone 16 from Fig. 2, also Col. 4, lines 26-35).

However, it is noted that Marshall, as modified by Boesen, doesn't disclose a speech recognition engine recognizing speech based on the microphone signal and the sensor signal. Conversely, Holzrichter teaches a speech recognition algorithm to be applied

after determining by a combiner if speech is present (speech recognition algorithm 68, combiner 67, and feature vectors 65 and 61 that represent the microphone and sensor signals, respectively, in Fig. 12, also Col. 26 lines 13-16, *the two feature vectors are further processed and combined and if the result is speech recognition, a speech recognition algorithm 68 is applied*).

It would have been obvious to one having ordinary skill in the art at the time the invention was made to have used the feature of and an in-ear speech sensor configured to sense a physical characteristic indicative of speech and output a sensor signal indicative of the sensed physical characteristic as taught by Boesen for Marshall's system because Boesen provides a bone conduction voice transmission apparatus and system that transmits voice sound using bone conduction and air conduction to obtain a pure voice sound signal for transmission minimizing interference from the surrounding sound environment (Col. 1, lines 6-12).

It would also have been obvious to one having ordinary skill in the art to use the feature of a speech recognition algorithm as taught by Holzrichter for Marshall's speech recognition system, as modified by Boesen, because Holzrichter provides the use of non-acoustic information in combination with acoustic information for speech recognition and related speech technologies.

As per claim 38, Marshall, in view of Boesen and Holzrichter, teach the speech recognition system of claim 37, further comprising:

a speech detector component configured to receive the sensor signal and output a speech detection signal indicative of whether the user is speaking based on the

sensor signal (Holzrichter: Col. 26, lines 34-38, where combiner 67 in Fig. 12 contains *algorithmic decision tree joining one nonacoustic speech recognition (NASR) and one conventional acoustic speech recognition (CASR) algorithm*; also Col. 25 line 63 to Col. 26 line 16 describes the process signals undergo until formed into feature vectors 61 from NASR and 65 from CASR (in Fig. 12), and further processed and combined to determine if speech is present in order to apply a speech recognition algorithm).

It would also have been obvious to one having ordinary skill in the art to use the feature of a speech detector component as taught by Holzrichter for Marshall's speech recognition system, as modified by Boesen, because Holzrichter provides the use of non-acoustic information in combination with acoustic information for speech recognition and related speech technologies.

As per claim 39, Marshall, in view of Boesen and Holzrichter, teach the speech recognition system of claim 38, further comprising:

a background speech removal component providing a modified speech signal based on the speech detection signal and the microphone signal (Holzrichter: processor 66 in Fig. 12, Col. 26, lines 4-7, *processor can include gain setting, speaker normalization, time adjustment, background removal, comparison to data from previous frames, and other well known procedures*).

It would also have been obvious to one having ordinary skill in the art to use the feature of a background speech removal component as taught by Holzrichter for Marshall's speech recognition system, as modified by Boesen, because Holzrichter

provides the use of non-acoustic information in combination with acoustic information for speech recognition and related speech technologies.

As per claim 40, Marshall, in view of Boesen and Holzrichter, teach the speech detection system of claim 39, wherein the speech recognition engine is configured to recognize speech represented by the modified speech signal (Holzrichter: speech recognition algorithm 68 in Fig. 12, Col. 26 lines 13-16, *the two feature vectors are further processed and combined and if the result is speech recognition, a speech recognition algorithm 68 is applied*).

It would also have been obvious to one having ordinary skill in the art to use the feature of the speech recognition engine configured to recognize speech represented by the modified speech signal, as taught by Holzrichter, for Marshall's speech recognition system, as modified by Boesen, because Holzrichter provides the use of non-acoustic information in combination with acoustic information for speech recognition and related speech technologies.

Conclusion

11. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

12. Helbing (US 2005/0038659) provides a method of operating a barge-in dialogue system, in which a speech recognition unit is not activated until a speech signal is detected by the speech activity detector (Paragraph [0034]).

13. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Natalie Lennox whose telephone number is (571) 270-1649. The examiner can normally be reached on Monday to Friday 9:30 am - 7 pm (EST).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571)272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

NL 08/16/2007

RICHEMOND DORVIL
SUPERVISORY PATENT EXAMINER